Explicable Question Answering

Endri Kacupaj

Smart Data Analytics, University of Bonn, Germany
kacupaj@iai.uni-bonn.de

Abstract. Question answering over Knowledge Graphs has emerged as an intuitive way of querying structured data sources and has witnessed significant progress over the years. However, there is still plenty of space for improvement and there exist specific challenges that are still far from being effectively solved. In this research project, we aim to address some of these challenges and provide innovative solutions in the field. Our research will mainly focus on deep learning approaches such as sequence to sequence models and ranking methods. We plan to contribute to the challenges of explicability and complex queries by further researching the areas and providing resources together with more robust models by using probabilistic methods and meta-learning approaches.

Keywords: Question Answering · Knowledge Graphs · Semantic Web · Machine Learning.

1 Introduction

Nowadays Question Answering (QA) systems over Knowledge Graphs (KG) have revolutionized the way of providing on-demand and accurate information. Most of the business sectors are moving towards and trying to adopt this type of systems - referred to also as chatbots - for providing more reliable services. The research area has already achieved significant contributions, by developing more and more robust methods by using different machine learning techniques.

2 State of the Art

The KGQA task involves answering a natural language question by using the information stored in a KG. The input question is first translated into a formal query language e.g., lambda calculus, lambda DCS, SPARQL. After, this formal query is executed over the KG to retrieve the answer. The task of converting a natural question into a formal KG query is called semantic parsing. The prediction models commonly used in semantic parsing can be categorized into - 1) classification, 2) ranking and 3) translation based models.

Classification models [18,20] are commonly used for simple queries. Simple queries comprise one subject entity, one relation, and one object entity. For these models it is assumed that formal queries follow a fixed structure. Therefore classification models do not preform well on complex queries. On the other hand, ranking models [2,8,17,3] that operate on a two-step procedure are considered better candidates for complex queries. The first step here is to find the top few probable query candidates, so that afterwards a neural network-based ranking model can be used to find the best candidate.

Finally, translation based models [13,15,7,5] treat semantic parsing as a translation problem. Usually, sequence to sequence models are used to translate an input natural language question to the corresponding logical form. The encoder is used to encode the input sequence and create context-dependent representations while the decoder generates the output sequence one token at a time, conditioning on the previously generated tokens and the input sequence. Different neural architectures are used for translation-based models such as RNN, CNN, and transformers.

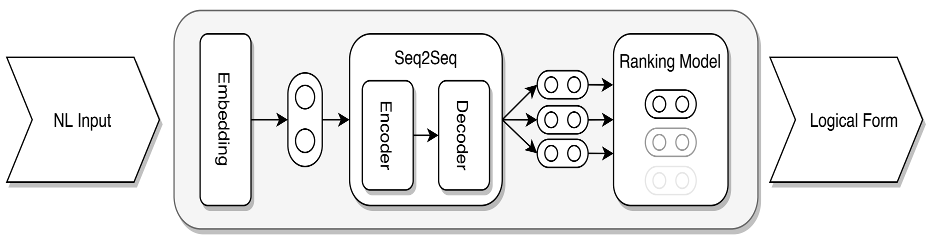

Fig. 1. A typical KGQA system, using a sequence to sequence and ranking model.

Figure 1 illustrates a typical way of implementing a KGQA system using sequence to sequence translation model alongside with a ranking model. The input to the system is a natural language question and the output is the best generated logical form. Initially, the system transforms the input question into embeddings and forwards it through an encoder-decoder model. There, by using approaches such as beam search, the model outputs multiple logical form candidates that might represent the correct interpretation of the question. After this step a ranking model is used for better matching the input with the respective logical form. The result is the final output of the system. This architecture serves as the core of our research and our work intends to improve the design and performance of such systems. We believe that there is optimization potential in at least two phases of this system. The first is the sequence to sequence model and the outputs it produces, and the second one is the ranking model.

3 Problem Statement and Contributions

Despite their current success, Question Answering over Knowledge Graphs (KGQA) systems still have many challenges to overcome. One of the biggest challenges is the question/query complexity [4,14]. Even though researchers have already done work in this field, today's systems are still suffering from handling multi-hop questions. Another challenge is the interoperability between different

KGs [4]. Most of KGQA systems are built on top of one particular knowledge graph (e.g. DBpedia, Wikidata, Freebase). It will be ideal if can have systems that can operate on different knowledge graphs simultaneously. This intends to make the system more robust and provide more concrete results. Furthermore, a huge challenge is multilinguality in those systems [4]. The ideal QA system should be able to accept inputs in different languages and still be able to provide a correct answer. This area is quite unexplored for KGQA and the main reason is the lack of resources. Currently, the available datasets [16] are quite small and not sufficient for deep learning approaches.

Another relevant challenge is the explicability of KGQA systems [14]. Currently, all QA systems are responsible for providing only the answer from the knowledge graph. But no hint is given of how the systems interpreted the input question and how it accomplished the way to the answer. Explicability in those systems intends to provide further details on what the answers means. Finally, another challenge is the robustness of KGQA systems [4]. While researchers advance the field regularly by providing new architectures that always push the state-of-the-art, we argue that not all the systems are appropriate for productive deployment. The robustness of those systems is something that definitely needs more attention.

The two important challenges that we aim to address are the **explicability**, and **question/query complexity**. Our goal is to use - mainly - machine learning techniques. Specifically, we intend to use deep learning approaches such as encoder-decoder and ranking models.

3.1 Explicability

In the context of explicability, we plan to focus on enhancing the interaction of the user with the QA system by providing additional information that helps to understand how the question was interpreted and the answer was retrieved.

In an attempt to enable the users to verify the answer provided by a QA system, researchers employ various techniques such as (i) revealing the generated formal query [11], (ii) graphical visualizations of the formal query [24] and (iii) verbalizing the formal query [19,10,6].

We take a different approach to addressing the problem of validating the answers given by the QA system. We aim to verbalize the answer in a way that it conveys not only the information requested by the user but also includes additional characteristics that are indicative of how the answer was determined. In order to do this we have built a dataset and experiment with different models that can perform reasoning. The dataset aims to cover the verbalizations of answers in a KGQA task. The goal here is to further support the answers of a QA system by providing a sentence verbalization that better captures the semantic content of the questions combined with the answer. This allows the users to better understand how the system interpreted their question and to better understand the meaning of the presented answer. Considering the question "Which Nobel has Marie Curie won?", a normal system would provide the answers [Nobel Prize in Chemistry, Nobel Prize in Physics]. It would be more informative if we can provide a verbalized answer stating "Marie Curie has been awarded Nobel prize both in Chemistry and Physics".

Alongside with the dataset, we intend to build models for handling this task. Initially, we plan to build models that can easily extend different existing QA systems for supporting the verbalization of answer. Our optimal goal here is to build an end-to-end QA system that performs both query construction and verbalization at once and, if possible, in one model. Most of the research here will be facilitating translations based systems.

3.2 Question/Query Complexity

Building QA systems for handling questions with high complexity is currently an active research field. By complex queries (also referred to as multi-hop questions), we consider questions with multiple entities and/or multiple relations that might require aggregations, filtering or ranking. These types of questions are quite difficult for the state-of-the-art systems. An example can be "Which awards do Marie Curie and Pierre Curie have in common?". The question here is still a simple example but is considered as a multi-hop question, since it requires 2 entities and one property to retrieve the answer.

Our work on question complexity is based on developing frameworks including grammars, deep semantic parsers, and models that will be able to capture the entire information of the input.

Considering the two KGQA challenges we believe that we can make new contribution by using probabilistic methods and meta-learning approaches. Probabilistic methods will help us handle the uncertainty in machine learning models and therefore will be a great basis for advancing the explicability of a KGQA system.

Moreover, meta-learning has been proposed as a framework to address the challenging few-shot learning setting. The key idea is to leverage a large number of similar few-shot tasks in order to learn how to adapt a base-learner to a new task, for which only a few labeled samples are available. Currently most meta-learning approaches (related to QA) refer to semantic parsing [1,12,22] and not directly to KGQA. We aim to research meta-learning approaches so that we can use them for developing more robust KGQA systems.

3.3 Research Questions

In order to investigate these two challenges, we aim to address the following main research question, which represents the philosophy behind the project:

- How can we enhance Knowledge Graph Question Answering systems to support explicability and complex queries?

In order to restrict and specify the scope of our research work, we plan to investigate the following sub-questions:

- How can translation and ranking models be incorporated in order to develop explicable KGQA systems?

– How we leverage probabilistic methods for improving translation based KGQA models to support a more expressive query language and answer verbalization?
– How can meta-learning be used for KGQA models regarding the explicability and complex queries?

Considering the research sub-questions, we aim to make the following contributions. We will develop systems using sequence to sequence and ranking methods for handling the explicability of a KGQA system and generate verbalizations that will surpass the state-of-the-art. We will research probabilistic methods in order to advance both sequence to sequence and ranking models for generating better verbalizations and handling complex queries. Finally, we will adopt meta-learning approaches to KGQA models, in particular, to target the explicability of the system and complex queries.

4 Research Methodology

During our research, we aim to focus on ranking and translation based models in order to address the challenges we mentioned. Translation models can be considered appropriate for generating the most probable candidates, while ranking models will be used to select the top candidate. We intend to contribute to both models using probabilistic and meta-learning approaches.

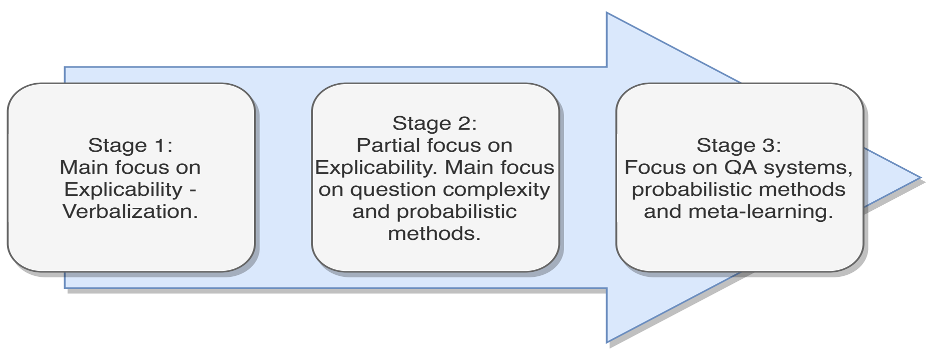

Fig. 2. Research project stages.

For addressing the first research question we plan to investigate approaches for generating multiple candidates using translations models (e.g. with beam search) and aim to further improve them alongside the ranking models. For the second question, we will further research probabilistic methods for generating more suitable candidates and also advancing the ranking models. Finally, we are interested in meta-learning techniques and will examine how to involve them for the KGQA task. The idea here is to create models that can be trained for one task using meta-learning approaches and allow them to be used more efficiently for other domains or tasks.

Figure 2 illustrates the stages of our research work. Initially we plan to focus on the explicability and partially on question complexity. In the second stage we plan to continue working on explicability but our main focus will turn to QA systems and to handling complex queries by using probabilistic methods. Finally, for the last stage, we aim to incorporate meta-learning approaches.

5 Evaluation Plan

Our evaluation plan focuses on the two KGQA challenges - explicability and complex queries. At the moment we do not intend to perform any specific evaluation for the probabilistic methods and meta-learning approaches. Our results will be evaluated by adopting commonly used metrics and will be compared to other state-of-the-art approaches for the respective tasks.

For the explicability, we have created a dataset named VQuAnDa with KGQA answer verbalizations. The dataset will aim to support both explicability and question answering task. Our goal will be to develop various models that can provide better verbalizations. The models will be evaluated using metrics such as the BLEU score or accuracy.

For evaluating the KGQA models, which we plan to develop for handling complex queries, we intend to use different available datasets such as LC-QuAD 1, 2 [23,9] and CSQA [21]. The whole evaluation process will be automated and no manual effort will be required. Our models are expected to surpass the state-of-the-art when applied to the upon mentioned datasets.

6 Preliminary Results

During the last months, we focused on answer verbalization for the KGQA task. The dataset called VQuAnDa[1] (Verbalization Question Answering Dataset) was built on top of the LC-QuAD dataset and intends to further support the answer verbalization of the respective questions. Currently the dataset contains natural language questions, with their respective SPARQL query and the verbalized answer. Both questions and answers are only in English language. To generate the dataset we followed a semi-automated approach. Initially, we retrieved the answers to all questions by using the DBpedia endpoint. Next, we generated the templates for the verbalized answers and we filled them with entities, properties, and query results. The last 2 steps had to be done manually in order to ensure the correctness of the verbalizations. First, we corrected and, if necessary, rephrased all answers to sound more natural and fluent. Finally, to ensure the grammatical correctness of the dataset, we peer-reviewed all the generated results. We plan to focus on this dataset and develop models that will be able to provide better and more accurate verbalizations. We also aim to extend the dataset with multiple verbalizations for each question. The work-related to the VQUAnDa dataset aims to cover the explicability part of our research.

[1] http://vquanda.sda.tech/

At the same time, we have worked on creating a QA framework for supporting complex queries on a scientific knowledge graph. The framework consists of different steps and models, which analyze the input natural language question and construct the respective query representation. Our framework is still in an early stage of development and we plan to further focus our work on it. Currently, the framework will be considered only for domain specific knowledge graphs but we plan to find ways and expand it in order to address open domain. For evaluating the system we built a dataset with complex queries that can be answered by the scientific knowledge graph.

Currently, we are working on releasing the first version of VQuAnDa dataset together with some sequence to sequence baseline models. We are already working on new models that can outperform the baselines and we expect to publish them later this year.

7 Conclusions

Our research focuses on Knowledge Graph Questions Answering systems. In particular, we emphasize on 2 main challenges related to the task. In the first one – explicability, we focus on how to provide better and more explicable results to the users. We handle explicability by focusing on answer verbalizations. The second challenge, which has become quite popular nowadays, is the one on complex queries. We intend to produce models and frameworks that can handle questions with high complexity (multi-hop with many entities and/or properties). At the same time, we aim to enhance current models by using probabilistic methods and meta-learning approaches.

References

1. Agarwal, R., Liang, C., Schuurmans, D., Norouzi, M.: Learning to Generalize from Sparse and Underspecified Rewards. arXiv e-prints (2019)
2. Bordes, A., Chopra, S., Weston, J.: Question answering with subgraph embeddings. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Doha, Qatar (2014)
3. Bordes, A., Usunier, N., Chopra, S., Weston, J.: Large-scale Simple Question Answering with Memory Networks. arXiv e-prints (2015)
4. Chakraborty, N., Lukovnikov, D., Maheshwari, G., Trivedi, P., Lehmann, J., Fischer, A.: Introduction to Neural Network based Approaches for Question Answering over Knowledge Graphs. arXiv e-prints (2019)
5. Cheng, J., Lapata, M.: Weakly-supervised Neural Semantic Parsing with a Generative Ranker. arXiv e-prints (2018)
6. Diefenbach, D., Dridi, Y., Singh, K., Maret, P.: Sparqltouser: Did the question answering system understand me? In: ISWC 2017 (2017)
7. Dong, L., Lapata, M.: Coarse-to-Fine Decoding for Neural Semantic Parsing. arXiv e-prints (2018)
8. Dong, L., Wei, F., Zhou, M., Xu, K.: Question answering over Freebase with multi-column convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint

Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Beijing, China (2015)
9. Dubey, M., Banerjee, D., Abdelkawi, A., Lehmann, J.: Lc-quad 2.0: A large dataset for complex question answering over wikidata and dbpedia. In: Proceedings of the 18th International Semantic Web Conference (ISWC). Springer (2017)
10. Ell, B., Harth, A., Simperl, E.: Sparql query verbalization for explaining semantic search engine queries. In: European Semantic Web Conference (2017)
11. Ferré, S.: Sparklis: an expressive query builder for sparql endpoints with guidance in natural language. Semantic Web (2017)
12. Guo, D., Tang, D., Duan, N., Zhou, M., Yin, J.: Coupling Retrieval and Meta-Learning for Context-Dependent Semantic Parsing. arXiv e-prints (2019)
13. He, X., Golub, D.: Character-level question answering with attention. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Austin, Texas (2016)
14. Kacupaj, E., Zafar, H., Lehmann, J., Maleshkova, M.: VQuAnDa: Verbalization QUestion ANswering DAtaset. e-prints (2019)
15. Liang, C., Berant, J., Le, Q., Forbus, K.D., Lao, N.: Neural Symbolic Machines: Learning Semantic Parsers on Freebase with Weak Supervision. arXiv e-prints (2016)
16. Lopez, V., Unger, C., Cimiano, P., Motta, E.: Evaluating question answering over linked data. Web Semantics Science Services And Agents On The World Wide Web (2013). https://doi.org/10.1016/j.websem.2013.05.006
17. Lukovnikov, D., Fischer, A., Lehmann, J., Auer, S.: Neural network-based question answering over knowledge graphs on word and character level. In: Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2017)
18. Mohammed, S., Shi, P., Lin, J.: Strong baselines for simple question answering over knowledge graphs with and without neural networks. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). Association for Computational Linguistics, New Orleans, Louisiana (2018)
19. Ngonga Ngomo, A.C., Bühmann, L., Unger, C., Lehmann, J., Gerber, D.: Sparql2nl: verbalizing sparql queries. In: Proceedings of the 22nd International Conference on World Wide Web (2013)
20. Petrochuk, M., Zettlemoyer, L.: SimpleQuestions nearly solved: A new upperbound and baseline approach. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium (2018)
21. Saha, A., Pahuja, V., Khapra, M.M., Sankaranarayanan, K., Chandar, S.: Complex Sequential Question Answering: Towards Learning to Converse Over Linked Question Answer Pairs with a Knowledge Graph. arXiv e-prints (2018)
22. Sun, Y., Tang, D., Duan, N., Gong, Y., Feng, X., Qin, B., Jiang, D.: Neural Semantic Parsing in Low-Resource Settings with Back-Translation and Meta-Learning. arXiv e-prints (2019)
23. Trivedi, P., Maheshwari, G., Dubey, M., Lehmann, J.: Lc-quad: A corpus for complex question answering over knowledge graphs. In: Proceedings of the 16th International Semantic Web Conference (ISWC). Springer (2017)
24. Zheng, W., Cheng, H., Zou, L., Yu, J.X., Zhao, K.: Natural language question/answering: Let users talk with the knowledge graph. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (2017)